biomolecules

MDPI

Article

The Budapest Amyloid Predictor and Its Applications

László Keresztes [1,†], Evelin Szögi [1,†], Bálint Varga [1], Viktor Farkas [2], András Perczel [2,3] and Vince Grolmusz [1,4,*]

[1] PIT Bioinformatics Group, Eötvös University, H-1117 Budapest, Hungary; keresztes@pitgroup.org (L.K.); szogi@pitgroup.org (E.S.); balorkany@pitgroup.org (B.V.)
[2] MTA-ELTE Protein Modeling Research Group, H-1117 Budapest, Hungary; farkasv@caesar.elte.hu (V.F.); perczel@chem.elte.hu (A.P.)
[3] Laboratory of Structural Chemistry and Biology, Eötvös University, H-1117 Budapest, Hungary
[4] Uratim Ltd., H-1118 Budapest, Hungary
* Correspondence: grolmusz@pitgroup.org
† Joint first authors.

Abstract: The amyloid state of proteins is widely studied with relevance to neurology, biochemistry, and biotechnology. In contrast with nearly amorphous aggregation, the amyloid state has a well-defined structure, consisting of parallel and antiparallel β-sheets in a periodically repeated formation. The understanding of the amyloid state is growing with the development of novel molecular imaging tools, like cryogenic electron microscopy. Sequence-based amyloid predictors were developed, mainly using artificial neural networks (ANNs) as the underlying computational technique. From a good neural-network-based predictor, it is a very difficult task to identify the attributes of the input amino acid sequence, which imply the decision of the network. Here, we present a linear Support Vector Machine (SVM)-based predictor for hexapeptides with correctness higher than 84%, i.e., it is at least as good as the best published ANN-based tools. Unlike artificial neural networks, the decisions of the linear SVMs are much easier to analyze and, from a good predictor, we can infer rich biochemical knowledge. In the Budapest Amyloid Predictor webserver the user needs to input a hexapeptide, and the server outputs a prediction for the input plus the $6 \times 19 = 114$ distance-1 neighbors of the input hexapeptide.

Keywords: amyloid; support vector machines; site-specific amyloidogenecity; Budapest Amyloid Predictor

Citation: Keresztes, L.; Szögi, E.; Varga, B.; Grolmusz, V.; Perczel, A.; Grolmusz, V. The Budapest Amyloid Predictor and Its Applications. *Biomolecules* **2021**, *11*, 500. https://doi.org/10.3390/biom11040500

Academic Editor: Martin Muschol

Received: 29 January 2021
Accepted: 23 March 2021
Published: 26 March 2021

1. Introduction

The primary structure of the proteins is characterized by their amino acid sequence. While the primary structure basically determines the spatial folding of the proteins, and, consequently, all chemical and biological properties of the given protein, inferring those properties from the amino acid sequence is a very difficult task. Here, we consider the amyloid predictors—tools, which tell us if a given amino acid sequence has or does not have the propensity to become amyloid.

Amyloids are misfolded protein aggregates [1,2], which—in contrast with the unstructured aggregates—have a well-defined structure, comprising parallel β-sheets [3,4]. Amyloids are present in numerous organisms in biology: for example, in healthy human pituitary secretory granules [5]; in the immune system of certain insects [6], the silkmoth chorion and some fish choria [7]; in human amyloidoses and several neurodegenerative diseases [8].

Most recently, on the analogy of the naturally occurring antiherpes activity of β-amyloids, synthetic amyloid peptides were developed, acting as amyloidogenic aggregation cores in certain viral proteins with high specificity [9]. This way, new amyloid-based antiviral pharmaceuticals can be developed in the very near future: the specific aggregation cores turn the viral proteins into insoluble amyloids. Consequently, potential amyloidogenecity may have direct pharmaceutical relevance.

Sequence-based amyloid predictors would help the understanding and the exploitation of the amyloid state of the proteins: instead of the difficult, costly, and slow wet-

laboratory tests, we can use the predictor on thousands or millions of inputs for enlightening the amyloidogenecity of proteins. A very recent review [10] covers the sequence-based amyloid-predictors, applying different strategies like AGGRESCAN [11], Zyggregator [12], netCSSP [13], and APPNN [14], among others.

In the last several years, the six-amino-acid-long peptides have become a model of studying amyloid formation [15–17]. The reason for this is twofold: first, numerous evidence shows the biological relevance of amyloid-forming hexapeptides [15,18–21]; second, one can form $20^6 = 64$ million hexapeptides from the 20 amino acids, which is a large—but not too large—and rich space of model molecules, whose structures are less complex and, therefore, easier to be dealt with as larger model spaces.

The APPNN predictor applies a machine-learning approach by training on 296 hexapeptides, selected from various sources, then predicts if a given hexapeptide is amyloidogenic or not. For longer sequences, it screens six-amino-acid-long sliding windows in longer polypeptide-chains to predict if they would form amyloid structures.

In this contribution, we construct and present a Support Vector Machine (SVM) predictor for hexapeptides, with better accuracy (84%) than most of the neural network-based tools (see [14] for a tabular comparison of the accuracy of those tools). We note that we do not repeat the comparative data described in [14], which evaluates numerous earlier published amyloidicity-prediction methods with APPNN. The main advantages of our new predictor, compared with other amyloid-predictors, are as follows:

(i) Simplicity: we used solely a linear SVM in its construction;
(ii) Transparency: no prefiltering and data manipulation were used in the construction of the predictor;
(iii) Truly experimental training set: The experimental hexapeptide Waltz database [15,16] was applied in the SVM training and data were not privately filtered, predicted, and constructed as in other predictors;
(iv) Free online availability, together with automatic prediction of the neighboring hexapeptides;
(v) Easy applicability for inferring location-dependent amyloidogenic properties of amino acids, as we describe below.

We also note that neural-network-based predictors are neither simple nor easy to apply, and inferring the causality of their classifications is a very difficult task. In the case of SVMs, especially for linear SVMs, the causality is much more transparent, as we demonstrate in Tables 1 and 2.

Table 1. The precomputed values from Equation (1) are listed in the rows, corresponding to the amino acids. The columns are correspond to their positions.

	1	**2**	**3**	**4**	**5**	**6**
A	−0.26	−0.32	−0.27	−0.14	−0.43	−0.22
R	−0.45	−0.41	−0.46	−0.33	−0.52	−0.35
N	−0.40	−0.34	−0.49	−0.27	−0.46	−0.30
D	−0.49	−0.43	−0.56	−0.41	−0.56	−0.36
C	−0.09	−0.21	0.03	−0.05	−0.17	−0.05
Q	−0.37	−0.30	−0.36	−0.34	−0.48	−0.32
E	−0.51	−0.41	−0.43	−0.30	−0.61	−0.39
G	−0.23	−0.37	−0.46	−0.37	−0.30	−0.33
H	−0.32	−0.26	−0.26	−0.30	−0.35	−0.25
I	−0.06	−0.08	0.26	0.09	−0.06	−0.07
L	−0.10	−0.18	0.02	0.04	−0.22	−0.13
K	−0.39	−0.45	−0.51	−0.35	−0.59	−0.32
M	−0.17	−0.25	−0.02	−0.10	−0.19	−0.18
F	−0.13	−0.11	0.05	−0.03	−0.13	−0.11
P	−0.56	−0.38	−0.56	−0.51	−0.42	−0.45

Table 1. *Cont.*

	1	2	3	4	5	6
S	−0.37	−0.35	−0.41	−0.30	−0.48	−0.23
T	−0.34	−0.33	−0.28	−0.23	−0.40	−0.23
W	−0.17	−0.17	−0.09	−0.06	−0.12	−0.16
Y	−0.23	−0.11	−0.13	−0.06	−0.18	−0.15
V	−0.05	−0.14	0.19	0.14	−0.19	0.01

Table 2. The amyloidogenecity order of the amino acids, decreasing from left to right.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	V	I	C	L	F	M	W	G	Y	A	H	T	S	Q	K	N	R	D	E	P
2	I	F	Y	V	W	L	C	M	H	Q	A	T	N	S	G	P	R	E	D	K
3	I	V	F	C	L	M	W	Y	H	A	T	Q	S	E	R	G	N	K	D	P
4	V	I	L	F	C	W	Y	M	A	T	N	H	E	S	R	Q	K	G	D	P
5	I	W	F	C	Y	M	V	L	G	H	T	P	A	N	Q	S	R	D	K	E
6	V	C	I	F	L	Y	W	M	A	T	S	H	N	Q	K	G	R	D	E	P

2. Methods

For the construction of the Budapest Amyloid Predictor, we have applied an artificial intelligence tool, the linear Support Vector Machine architecture [22], abbreviated as SVMs. In linear Support Vector Machines, $n + m$ data points correspond to $n + m$ vectors, each of k dimensions, $x_1, x_2, \ldots, x_n$, and $y_1, y_2, \ldots, y_m$, and the goal is to find a hyperplane that optimally separates the x and the y data points.

The mathematical foundation of SVMs is the trivial observation that any subset of the $k + 1$ vertices of a k-dimensional simplex can be separated by a hyperplane from its complementer: it is obvious in a case of a triangle (a 2-dimensional simplex) or a tetrahedron (a 3-dimensional simplex). The mathematical problem becomes more interesting if the data points are not in general positions, or the separation is done in a smaller dimensional Euclidean space than the number of data points [23].

Usually, the dataset is partitioned into a training and a testing subset: the first one is applied in the construction of the SVM, the second one is used for testing the resulting tool.

We have used the Waltz database [15,16] of 1415 hexapeptides, annotated to be amyloidogenic (514 peptides) or nonamyloidogenic (901 peptides). The annotation in the Waltz database was made by Thioflavin-T binding assays and literature search [15,16]; consequently, it is based on experimental evidence. Similarly, as in [14], two vectorial representations of the hexapeptides were considered in the present work. The first is the simple translation of the 20 amino acid names into vectors, each amino acid X was corresponded to a length-20 0-1 vector, with a single 1-coordinate identifying X (called orthogonal representation). This way, a hexapeptide is described by a $6 \times 20 = 120$-dimensional 0-1 vector.

The second one is based on the AAindex, a physicochemical property database of 553 properties [24]; https://www.genome.jp/aaindex/ (accessed on 25 March 2021). In this representation, each amino acid corresponds to a 553-dimensional vector, and each hexapeptide to a $6 \times 553 = 3318$-dimensional vector.

From the 1415 (514 amyloid-, 901 nonamyloid-) hexapeptides found in Waltz database, we selected 158 amyloid and 309 nonamyloid hexapeptides randomly for the test set (roughly 33%). We used the remaining hexapeptides for training our linear SVM. We used the sklearn LinearSVC object from the SciKit-learn Python library [25] for constructing the classifier.

The orthogonal representation yielded approximately 80% accuracy, while the AAindex-based produced a much better accuracy; because of this, we have chosen the second, AAindex-based representation in what follows.

The classifier simply computes the sign of the $w \cdot z + b$ values for the 3318-long z vectors, corresponding to a hexapeptide, where w is a 3318-dimensional weight vector and

b is a scalar; if this sign is positive, then the prediction is "amyloidogenic", otherwise, it is "nonamyloidogenic".

On the 467 test examples, we achieved 127 true positives, 31 false positives, 266 true negatives, and 43 false negatives. The resulting classifier's performance for unseen examples is 0.8415 ± 0.0331 with 95% confidence. Based on test performance: ACC = 0.84, TPR = 0.75, TNR = 0.9, PPV = 0.8, NPV = 0.86, (that is, accuracy, true positive ratio, true negative ratio, positive predictive value, negative predictive value, respectively). The accuracy of our SVM is better than or on par with that of APPNN [14].

Figure 1 shows the ROC (Receiver Operating Characteristics) curve of the Budapest Amyloid Predictor. The AUC (Area Under Curve) value is 0.89. The precision-recall curve is provided as Figure S1 in the supporting material.

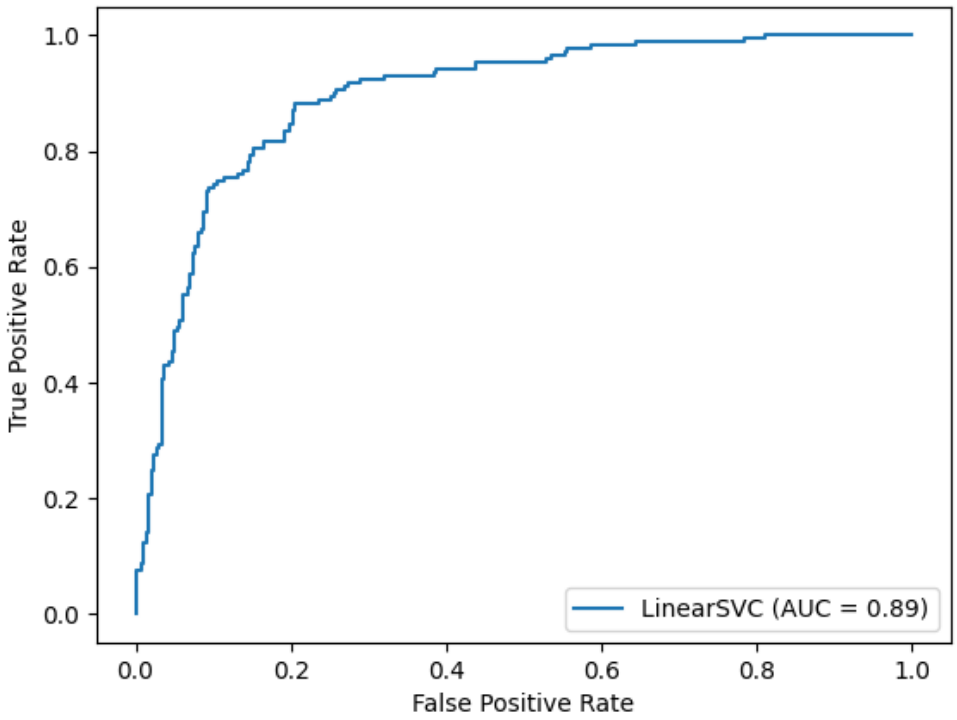

Figure 1. The ROC (Receiver Operating Characteristics) curve of the Budapest Amyloid Predictor. The AUC (Area Under Curve) value is 0.89. The precision-recall curve is provided as Figure S1 in the supporting material.

The Budapest Amyloid Predictor uses the above mentioned SVM. For verifying the underlying method on differently selected random training and test-sets, we used 10-fold cross validations, with the construction of 10 distinct SVMs, detailed in the online Supporting Material. The accuracies of the 10 SVM models were between 73% and 86% (see Table S1 in the supporting material). The analogs of our Table 1 for these SVMs are listed as supplementary Tables S2–S11.

3. Discussion and Results

The Budapest Amyloid Predictor webserver is available at the site https://pitgroup.org/bap/ (accessed on 25 March 2021). The user needs to input a hexapeptide with 6 capital letters, and the server returns the prediction for the query, plus the predictions of all 114 (= 6×19) 1-Hamming-distance neighbors of the query. If the hexapeptide is listed in the Waltz DB, then the "known" word appears next to prediction; otherwise, the "predicted" word appears.

Generally, it is very difficult to follow what a deep neural network does with a given input. In the case of a linear SVM, it is straightforward: for input vector z, the $w \cdot z + b$

quantity is computed, where the weight vector w and the scalar b are computed in the course of the SVM construction, and the sign of this quantity determines the prediction. Instead of specifying the 3318-dimensional weight vector w here, we present a very compact representation of the SVM in the next section. This representation not only specifies the predictor in a very simple way, but it also opens up novel insights of the amyloidogenic and nonamyloidogenic hexapeptides.

One of the greatest advantages of the linear SVM predictions is that we can easily see the reasons behind the decision of the model. If our model is accurate enough, then, from the coefficients of the normal vector of the separating hyperplane, the weight-differences of the distinct variables can be derived. We apply this observation below.

The following matrix enlightens the details of the decision of the linear SVM. Clearly, representing every amino acid by a 553-dimensional vector is highly redundant, since we have only 20 amino acids—that is, only 20 different 553-dimensional vectors exist in this representation. Therefore, we can write with $\ell = 553$:

$$w \cdot z = \sum_{i=1}^{6\ell} w_i z_i = \sum_{j=1}^{6} \sum_{i=(j-1)\ell+1}^{j\ell} w_i z_i. \quad (1)$$

For each fixed $j = 1, 2, \ldots, 6$, the $\ell = 553$ z_i's are determined by the jth amino acid of the hexapeptide, this way, all the possible $6x20 = 120$ second sums (for six positions and 20 amino acids) can be precomputed. Table 1 lists these precomputed values, the 6 values of j correspond to the columns and the amino acids to the rows.

The value of (1) can now be easily computed by adding exactly one item from each column, determined by the first, second, ..., sixth amino acid of the hexapeptide, plus the value of $b = 1.083$. For example, one can classify the hexapeptide AAEEAA by computing the sign of $(-0.26 - 0.32 - 0.43 - 0.30 - 0.43 - 0.22 + 1.083) = -0.88$, that is, -1, which predicts that AAEEAA is not amyloidogenic.

By observing Table 1, one can easily derive an amyloidogenecity order of the amino acids for each position from 1 through 6: we just sort the columns in decreasing order and substitute the amino acids in the rows of Table 2 (from left to right). For example, in the first column of Table 1, the largest number corresponds to V, the second largest to I, so the first element of the first row of Table 2 is V, the second is I, and so on.

In Table 2, the amyloidogenecity order decreases from left to right.

The hydrophobic amino acids valine, isoleucine, phenylalanine, tyrosine, and tryptophan populate the left portion of Table 2, these amino acids are naturally more probable to form amyloids. Interestingly, alanine is not in that region, while cysteine is there.

Aspartic acid, lysine, asparagine, and glutamic acid populate the right end. Naturally, proline, the "structure breaker", appears mostly at the right end, as one of the least amyloidogenic amino acids, but not in every row: in row 5, it is in position 12.

If there were no site-specific amyloidogenic properties of the amino acids, then all columns of Table 2 would be homogenic, i.e., every column would contain the same amino acid. Proline seems to be more "amyloid-breaker" in the ends and in the center of the hexapeptides, while much less so in the second and fifth position.

This table shows a remarkable difference in the amyloidogenecity order of the six positions of the hexapeptides: we believe that Table 2 is the most striking application of the Budapest Amyloid predictor.

Comparison with Earlier Work

The location-dependent amyloidogenic properties of amino acids in hexapeptides were studied earlier in a conference paper [26] and the position-specific amyloidogenic properties of the amino-acids were listed in a Table (Table 1 in [26]). We note that our Table 2 is substantially different from that list: we order the 20 amino acids in amyloidogenic order in each of the 6 positions of hexapeptides. Additionally, in Table 2, proline has a very distinct structure-breaker property, while no such observation was found in Table 1 of [26].

The work [26] applied 139 amyloid and 168 nonamyloid peptides (after a nondetailed filtering procedure) for statistical analysis of amino acid frequencies in the positions of hexapeptides (see Figure 2 and Table 1 in [26]). We used a much larger dataset (514 amyloidogenic and 901 nonamyloidogenic hexapeptides) and our method is not simple frequency analysis, but a much deeper artificial intelligence approach.

Using SVMs for amyloid prediction is not without precedence: In [27], a nonlinear SVM is constructed for hexapeptide amyloid prediction. Instead of using experimentally identified amyloidogenic hexapeptides, the authors of [27] constructed an in-house "Hexpepset dataset", where the positive and negative hexapeptides were gained from six amino acids sliding windows of known amyloid and nonamyloid proteins. We note that we use the experimentally verified Waltz dataset [15,16] of hexapeptides for training our linear SVM (each hexapeptide in the Waltz DB is annotated experimentally and not by theoretical inference with sliding windows of known amyloids or nonamyloids). Additionally, we attain a 84% accuracy on the experimental Waltz dataset, compared to the 81% accuracy of the much more complex, nonlinear SVM of [27] on a theoretically (sliding windows) and ad hoc constructed, nonexperimental, "Hexpepset dataset" of hexapeptides.

Supplementary Materials: The following are available at https://www.mdpi.com/2218-273X/11/4/500/s1, Figure S1: The precision-recall curve of the Budapest Amyloid Predictor, Table S1: The results of the 10-fold cross validations, Table S2: The analogue of Table 1 in the main text for cross validation Round 1, Table S3: the analogue of Table 1 in the main text for cross validation Round 2, Table S4: The analogue of Table 1 in the main text for cross validation Round 3, Table S5: The analogue of Table 1 in the main text for cross validation Round 4, Table S6: The analogue of Table 1 in the main text for cross validation Round 5, Table S7: The analogue of Table 1 in the main text for cross validation Round 6, Table S8: The analogue of Table 1 in the main text for cross validation Round 7, Table S9: The analogue of Table 1 in the main text for cross validation Round 8, Table S10: The analogue of Table 1 in the main text for cross validation Round 9, Table S11: The analogue of Table 1 in the main text for cross validation Round 10.

Author Contributions: A.P., V.F., and V.G. have initiated the study and evaluated results, L.K. and E.S. constructed the SVM for the prediction, B.V. constructed the webserver, V.G. has overseen the work and wrote the paper. A.P., V.F., and V.G. secured funding. All authors have read and agreed to the published version of the manuscript.

Funding: B.V. and V.G. were partially supported by the VEKOP-2.3.2-16-2017-00014 program, supported by the European Union and the State of Hungary, co-financed by the European Regional Development Fund, L.K., E.S. and V.G. by the NKFI-127909 L.K., E.S. and V.G. were supported in part by the EFOP-3.6.3-VEKOP-16-2017-00002 grant, supported by the European Union, co-financed by the European Social Fund. L.K., E.S., A.P., V.F. and V.G. were partially supported by the ELTE Thematic Excellence Programme (Szint+) supported by the Hungarian Ministry for Innovation and Technology. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Budapest Amyloid Predictor webserver is available freely at https://pitgroup.org/bap (accessed on 25 March 2021).

Conflicts of Interest: The authors declare no conflicting interests.

References

1. Horváth, D.; Menyhárd, D.K.; Perczel, A. Protein aggregation in a nutshell: The splendid molecular architecture of the dreaded amyloid fibrils. *Curr. Protein Pept. Sci.* **2019**, *20*, 1077–1088. [CrossRef] [PubMed]
2. Taricska, N.; Horvath, D.; Menyhard, D.K.; Akontz-Kiss, H.; So, M.; Goto, Y.; Fujiwara, T.; Perczel, A. The route from the folded to the amyloid state: Exploring the potential energy surface of a drug-like miniprotein. *Chem. Eur. J.* **2020**, *26*, 1968–1978. [CrossRef] [PubMed]
3. Takács, K.; Varga, B.; Grolmusz, V. PDB _Amyloid: An extended live amyloid structure list from the PDB. *FEBS Open Bio* **2019**, *9*, 185–190. [CrossRef]

4. Takacs, K.; Grolmusz, V. On the border of the amyloidogenic sequences: Prefix analysis of the parallel beta sheets in the PDB_Amyloid collection. *arXiv* **2020**, arXiv:2003.02942.
5. Maji, S.K.; Perrin, M.H.; Sawaya, M.R.; Jessberger, S.; Vadodaria, K.; Rissman, R.A.; Riek, R. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **2009**, *325*, 328–332. [CrossRef]
6. Falabella, P.; Riviello, L.; Pascale, M.; Di Lelio, I.; Tettamanti, G.; Grimaldi, A.; Pennacchio, F. Functional amyloids in insect immune response. *Insect Biochem. Mol. Biol.* **2012**, *42*, 203–211. [CrossRef]
7. Iconomidou, V.A.; Hamodrakas, S.J. Natural protective amyloids. *Curr. Protein Pept. Sci.* **2008**, *9*, 291–309. [CrossRef]
8. Soto, C.; Estrada, L.; Castilla, J. Amyloids, prions and the inherent infectious nature of misfolded protein aggregates. *Trends Biochem. Sci.* **2006**, *31*, 150–155. [CrossRef]
9. Emiel, M.; Kenny, R.; Rodrigo, G.; Ladan, K.; Laleh, K.; van der Kant, R.; Maxime, S.; Bert, H.; Meine, R.; Hannah, W.; et al. Reverse engineering synthetic antiviral amyloids. *Nat. Commun.* **2020**, *11*, 2832. [CrossRef]
10. Santos, J.; Pujols, J.; Pallarès, I.; Iglesias, V.; Ventura, S. Computational prediction of protein aggregation: Advances in proteomics, conformation-specific algorithms and biotechnological applications. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1403–1413. [CrossRef] [PubMed]
11. Conchillo-Sole, O.; de Groot, N.S.; Aviles, F.X.; Vendrell, J.; Daura, X.; Ventura, S. Aggrescan: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinform.* **2007**, *8*, 65. [CrossRef]
12. Gian Gaetano Tartaglia and Michele Vendruscolo. The zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **2008**, *37*, 1395–1401. [CrossRef]
13. Kim, C.; Choi, J.; Lee, S.J.; Welsh, W.J.; Yoon, S. Netcssp: web application for predicting chameleon sequences and amyloid fibril formation. *Nucleic Acids Res.* **2009**, *37*, W469–W473. [CrossRef]
14. Familia, C.; Dennison, S.R.; Quintas, A.; Phoenix, D.A. Prediction of peptide and protein propensity for amyloid formation. *PLoS ONE* **2015**, *10*, e0134679. [CrossRef] [PubMed]
15. Beerten, J.; Van Durme, J.; Gallardo, R.; Capriotti, E.; Serpell, L.; Rousseau, F.; Schymkowitz, J. WALTZ-DB: A benchmark database of amyloidogenic hexapeptides. *Bioinformatics* **2015**, *31*, 1698–1700. [CrossRef]
16. Louros, N.; Konstantoulea, K.; De Vleeschouwer, M.; Ramakers, M.; Schymkowitz, J.; Rousseau, F. WALTZ-DB 2.0: An updated database containing structural information of experimentally determined amyloid-forming peptides. *Nucleic Acids Res.* **2020**, *48*, D389–D393. [CrossRef] [PubMed]
17. Louros, N.; Orlando, G.; De Vleeschouwer, M.; Rousseau, F.; Schymkowitz, J. Structure-based machine-guided mapping of amyloid sequence space reveals uncharted sequence clusters with higher solubilities. *Nat. Commun.* **2020**, *11*, 3314. [CrossRef] [PubMed]
18. Hauser, C.A.; Deng, R.; Mishra, A.; Loo, Y.; Khoe, U.; Zhuang, F.; Hauser, U.A. Natural tri-to hexapeptides self-assemble in water to amyloid β-type fiber aggregates by unexpected α-helical intermediate structures. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1361–1366. [CrossRef] [PubMed]
19. Tenidis, K.; Waldner, M.; Bernhagen, J.; Fischle, W.; Bergmann, M.; Weber, M.; Kapurniotu, A. Identification of a penta-and hexapeptide of islet amyloid polypeptide (iapp) with amyloidogenic and cytotoxic properties. *J. Mol. Biol.* **2000**, *295*, 1055–1071. [CrossRef] [PubMed]
20. Reches, M.; Gazit, E. Amyloidogenic hexapeptide fragment of medin: homology to functional islet amyloid polypeptide fragments. *Amyloid Int. J. Exp. Clin. Investig. Off. J. Int. Soc. Amyloidosis* **2004**, *11*, 81–89. [CrossRef]
21. Iconomidou, V.A.; Chryssikos, G.D.; Gionis, V.; Galanis, A.S.; Cordopatis, P.; Hoenger, A.; Hamodrakas, S.J. Amyloid fibril formation propensity is inherent into the hexapeptide tandemly repeating sequence of the central domain of silkmoth chorion proteins of the a-family. *J. Struct. Biol.* **2006**, *156*, 480–488. [CrossRef] [PubMed]
22. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
23. Keresztes, L.; Szogi, E.; Varga, B.; Grolmusz, V. Identifying super-feminine, super-masculine and sex-defining connections in the human braingraph. *arXiv* **2019**, arXiv:1912.02291
24. Kawashima, S.; Pokarowski, P.; Pokarowska, M.; Kolinski, A.; Katayama, T.; Kanehisa, M. Aaindex: Amino acid index database, progress report 2008. *Nucleic Acids Res.* **2008**, *36*, D202–D205. [CrossRef]
25. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
26. Thangakani, A.M.; Kumar, S.; Velmurugan, D.; Gromiha, M.M. Distinct position-specific sequence features of hexa-peptides that form amyloid-fibrils: Application to discriminate between amyloid fibril and amorphous ?-aggregate forming peptide sequences. *BMC Bioinform.* **2013**, *14* (Suppl. 8), S6. [CrossRef] [PubMed]
27. Tian, J.; Wu, N.; Guo, J.; Fan, Y. Prediction of amyloid fibril-forming segments based on a support vector machine. *BMC Bioinform.* **2009**, *10*, S45. [CrossRef]